- A reminder: Please submit this and every homework as a single PDF document.

- As always, be sure to download fresh copies of the data sets for this assignment before beginning.

- Due Date: Friday, April 16 at 11:59pm.

Part 1: Barley data.

Immer et al. (1934) conducted an experiment to compare the yield of four varieties of barley (manchuria, velvet, trebi and peatland) when paired with two different fertilizers ("A" and "B"). Four locations in Minnesota were selected for the experiment. At each location, 8 plots were established, and the variety × fertilizer combinations were randomized to the plots such that each unique variety × fertilizer combination appeared exactly once at each location. The yield of each plot was measured at the end of the growing season. The data are contained in the file `barley.txt`, which can be found on the course website. The data set contains the following variables:

loc The [loc]ation: St. Paul, Waseca, Morris, or Duluth

var The [var]iety of barley

fertilizer The type of fertilizer provided

yield The response.

The investigator is interested in characterizing the differences among the barley varieties, between the fertilizer treatments, and any possible interaction between barley variety and fertilizer. The location is considered a blocking factor. The investigator is interested in drawing inferences that apply to these four specific locations, suggesting that the blocking factor should be modeled using a fixed effect. Use PROC GLM to answer question 1, and also answer question 2.

Part 2: Watermelon data

An experiment compared the effects of two different fertilizer application methods (broadcast vs. band) and two different fertilizer application rates (low vs. high) on watermelon yield. Application methods and application rates were crossed to form four unique treatment combinations. Treatment combinations were assigned to a total of sixteen (16) experimental units arranged in a Latin-squares design. You can find the data in the file `melon.txt` on the course website. The data set contains the following variables:

rblock The row block.

cblock The column block.

method The method of fertilizer application.

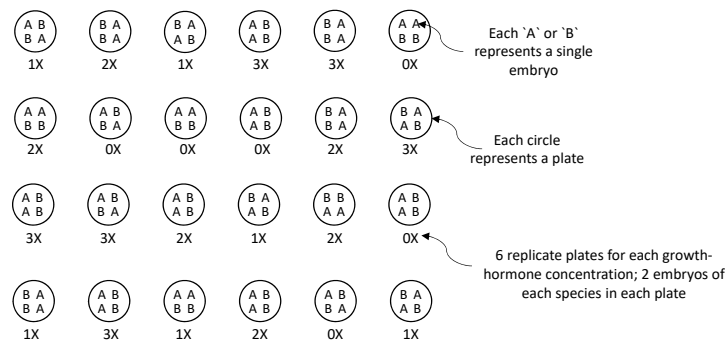**rate** The rate of fertilizer application.

yield The response.

Jan proposes to fit an ANOVA model to these data that includes the following terms:

- Full factorial effects (main effects and interaction) for the `method` and `rate` factors.

- A fixed effect for the row block.

- A fixed effect for the column block.

- Residual error.

Answer question 3 on the basis of this proposed model. (You do not actually have to do any model fitting for this question, although you are welcome to do so if you wish.)

Part 3: Effect of growth hormone concentration on pine seedlings

A plant biologist is interested in characterizing the effects of 4 concentrations of a growth hormone (0, 1x, 2x, 3x) on the growth of seedlings of two pine species (A and B). Species A and B were of specific interest to the researcher (A was known to be tolerant to acid rain whereas B was not tolerant). In this experiment, 24 petri dishes (also called plates) were prepared, 6 for each of the 4 growth hormone concentrations. Each plate contained 2 embryos from each species, for a total of 4 embryos per plate. The 24 plates were placed in randomly assigned locations in an incubator, and after 2 weeks the length of the primary shoot was measured for every embryo in every plate. The data set thus includes 96 records (24 plates × 4 embryos per plate.) The response of interest is the length of the primary shoot. A schematic of the design is shown below.

The data are contained in the file `pine.txt`, which can be found on the course website. The data set contains the following variables:

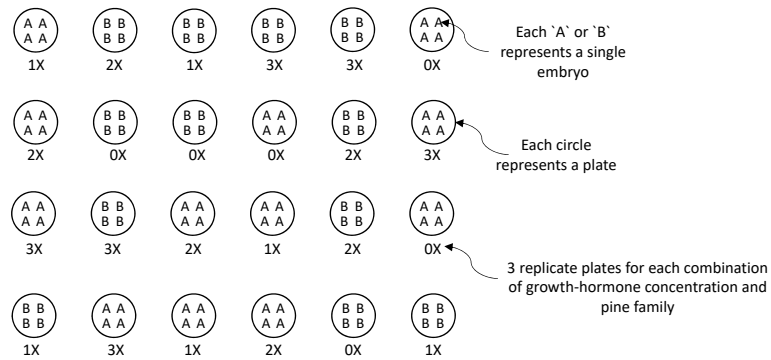**enzyme** The enzyme concentration of the plate

plate A variable to distinguish the replicate plates assigned to each enzyme concentration

species The species of the seedling

logy The log of the length of the primary shoot, the response.

Use these data to answer questions 4 – 7.

Here is an alternative version of the same experiment. This experiment also investigates the effects of 4 concentrations of a growth hormone (0, 1x, 2x, 3x) on the growth of seedlings of two pine species (A and B). Species A and B are again of specific interest to the researcher. This experiment also involved 24 petri dishes. However, in this experiment, 3 plates were assigned to each of the 8 unique combinations of growth-hormone concentration and pine species. Each plate contained 4 embryos, all of the same species. This data set also includes 96 records (24 plates × 4 embryos per plate.) The response of interest is the length of the primary shoot. A schematic of this modified design is shown below.

(A A / A A)	(B B / B B)	(B B / B B)	(B B / B B)	(B B / B B)	(A A / A A)	Each `A` or `B` represents a single embryo
1X	2X	1X	3X	3X	0X	
(A A / A A)	(B B / B B)	(B B / B B)	(A A / A A)	(B B / B B)	(A A / A A)	Each circle represents a plate
2X	0X	0X	0X	2X	3X	
(A A / A A)	(B B / B B)	(A A / A A)	(A A / A A)	(B B / B B)	(A A / A A)	3 replicate plates for each combination of growth-hormone concentration and pine family
3X	3X	2X	1X	2X	0X	
(B B / B B)	(A A / A A)	(A A / A A)	(A A / A A)	(B B / B B)	(B B / B B)	
1X	3X	1X	2X	0X	1X	

Answer question 8 based on this design. (There are no data available for this alternative design; your task is instead to contrast the two designs.)

ST 512 Homework 6 Questions.

1. (6 points) Provide a brief verbal answer to each question below. Support your answer with an appropriate test statistic and p-value. Show some work or computer code. 2 points each.

    (a) Is there evidence that the effect of the fertilizer treatment differs among the barley varieties?

    (b) Is there evidence that the mean yield differs among the barley varieties?

    (c) Is there evidence that the yield differs between the two fertilizers?

2. (2 points) A second analyst proposes analyzing the barley data with location treated as a third experimental factor. In other words, the analyst proposes to fit the model with full factorial effects for variety, fertilizer, and location. (By "full factorial effects", we mean all main effects, all first-order interactions, and the second-order interaction.) Is it reasonable to fit such a model to these data? Answer 'yes' or 'no', and briefly defend your answer.

3. (2 points) Complete the degrees of freedom accounting for the ANOVA model proposed for the melon data.

Source	df
Method	
Rate	
Method $*$ Rate	
Row block	
Column block	
Residual error	
Total	

4. (2 points) Identify the experimental unit for the hormone concentration treatment and for the pine species.

    • Experimental unit for hormone-concentration treatment:

    • Experimental unit for pine-species treatment:

5. (2 points) Is there evidence that the difference between the two species depends on the concentration of growth hormone? Provide a brief verbal answer, and support your answer with an appropriate test statistic and p-value. Show some work or computer code.

6. (2 points) Here is an equation for the usual split-plot model for these data:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + P_{k(i)} + \varepsilon_{ijkl}$$

where

    • $i = 1, 2, 3, 4$ is an index for the growth-hormone treatment

    • $j = 1, 2$ is an index for the species

    • $k = 1, 2, \ldots, 6$ is an index for the replicate plates assigned to each level of the growth hormone

- $l = 1, 2$ is an index for the replicate seedling of each species within each plate
- μ is a reference level
- α_i's are the (fixed) effect parameters that capture differences among the levels of the growth-hormone treatment
- β_j's are the (fixed) effect parameters that capture differences between the species
- $(\alpha\beta)_{ij}$'s are the (fixed) effect parameters that capture the hormone-by-species interaction
- $P_{k(i)}$ is a random effect for the replicate plates, $P_{k(i)} \sim \mathcal{N}\left(0, \sigma_P^2\right)$
- ε_{ijkl} is residual (split-plot) error, $\varepsilon_{ijkl} \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$

Based on your model fit, give estimates of the variances of both random effects. That is, give estimates of σ_P^2 and σ_ε^2.

7. (2 points) Prior to conducting the experiment, the investigator determined that she was especially interested in comparing the growth rates of the two species when the concentration of the growth hormone is 0x. Is there evidence that the growth rates of the two species differ when the concentration of the growth hormone is 0x? Provide a brief verbal answer, and support your answer with an appropriate statistical procedure. Show some work or computer code.

8. (2 points) For the alternative design of the pine seedling study, identify the experimental unit for the hormone concentration treatment and for the pine species.

- Experimental unit for hormone-concentration treatment:
- Experimental unit for pine-species treatment: